# Mobile Robotic Platform for Multi-Disease Retinal  Screening: A Comprehensive Review

*Digvijay J. Pawar[1], Shashi Kant Gupta[2]*

[1]Postdoctoral Researcher, Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia;

[2]Lincoln University College, Petaling Jaya, Selangor Darul Ehsan-47301, Malaysia

Email ID - [1]pdf.Digvijay@lincoln.edu.my, [2]raj2008enator@gmail.com, [2]shashigupta@lincoln.edu.my

**Abstract:** Retinal diseases such as diabetic retinopathy, age-related macular degeneration, glaucoma, cataract-related retinal changes and retinal vascular disorders remain major causes of preventable vision loss. However this problem is more serious in rural and resource constrained regions because screening still depends on trained specialists, manual fundus interpretation and hospital-based imaging systems. Recent studies show that deep learning, vision transformers, hybrid CNN-transformer models and multimodal systems can classify retinal diseases with high accuracy. Yet, many models are designed for offline testing or single-device diagnosis. They do not fully solve the field-level screening problem where tasks like image capture, patient alignment, local inference and remote consultation must work together. Therefore, this review suggests the need for an autonomous robotic retinal screening system which combines non-mydriatic fundus imaging, edge-AI inference and tele-ophthalmology support. The proposed framework uses CNN modules for local lesion features, transformer modules for global retinal context and optional multimodal fusion with physiological inputs. Model compression is done through quantization and pruning is considered for real-time edge deployment. Thus, the review connects recent retinal AI advances with the practical engineering pathway for scalable multi-disease screening in clinics, outreach camps and underserved healthcare settings.

**Keywords**: multi-disease retinal screening; fundus imaging; edge AI; tele-ophthalmology; mobile robots

## Introduction

Retinal diseases are one of the major causes for vision loss worldwide, and many of them usually progress silently in early stages, which leads to delayed diagnosis and permanent retinal damage. Although the retinal screening currently depends on fundus imaging, OCT imaging, and manual interpretation by trained ophthalmologists, this particular approach is very difficult to implement in resource-limited regions such as rural areas. In rural areas, the specialists, imaging facilities, and continuous screening support is quite limited. Recently, strong progress is seen in AI-based solutions in the field of multi-disease retinal detection, which includes ultra-wide fundus image classification by DUAN and Tu with an accuracy of above 0.980 for common diseases[1], retinal vision transformer-based multi-label classification by Wang et al. [2], and CAD-EYE approach which uses MobileNet and EfficientNet feature fusion with 98% accuracy on 65,871 fundus images [3]. However, in the most existing systems focus is mainly  on image classification and this does not address the complete screening workflow. Therefore, in this review we have considered autonomous multi-disease retinal screening as a system-level engineering problem by combining a mobile robotic platform, non-mydriatic fundus imaging, CNN-transformer based feature extraction, edge-AI inference, model compression and tele-ophthalmology support to develop a scalable screening framework for clinics, outreach camps and resource-limited healthcare settings.

**Related work**

Early retinal image analysis depends on manual grading and classical image-processing methods. These methods rely on the features such as optic disc shape, blood vessels, hemorrhages, exudates and macular changes. However, these methods were less reliable under varying image quality and in cases with multiple diseases. CNNs have improved the retinal disease classification by automatically learning lesion-level features from fundus and OCT images, however in many studies remained limited to diabetic retinopathy or binary detection. Therefore, recent research has shifted towards multi-class, multi-label, transformer-based and multimodal retinal screening.

Recent works have shown strong progress in this direction. Wang et al. proposed a retinal vision transformer for multi-label retinal disease classification and reported better performance than CNN models such as ResNet, VGG, DenseNet and MobileNet on the ODIR-2019 dataset [1]. Khalid et al. developed CAD-EYE, where MobileNet and EfficientNet features were fused for classifying diabetic retinopathy, hypertensive retinopathy, glaucoma and contrast-related eye diseases. The model was trained using 65,871 fundus images and reported 98% classification accuracy [2]. Duan and Tu used 10,612 ultra-widefield fundus images covering sixteen fundus diseases and reported accuracy above 0.980 for common diseases such as retinal vein occlusion, age-related macular degeneration and diabetic retinopathy [3]. These works prove that multi-disease retinal screening is technically feasible when large and diverse datasets are available.

Hybrid and multimodal approaches are also becoming important. Zedadra et al. proposed a multimodal AI framework using fundus images, OCT images and clinical information through CNN, GNN and LLM components. The model reported strong performance on RetinalOCT and RFMiD datasets [4]. Sevik and Mutlu developed a unified deep learning framework for nine retinal pathology classes using ResNet-152, EfficientNetV2 and a YOLOv11-based classifier [5]. Kulkarni and Reddy combined Intrinsic Mode Function filtering with LightCNN and achieved high accuracy on DIARETDB fundus datasets, showing the value of lightweight models for screening [6]. Rahman et al. proposed OpthaNet for cataract, diabetic retinopathy and glaucoma classification using attention-enhanced EfficientNetB3, MobileNetV2 and ViT models [7]. Yusufoğlu et al. proposed DenseViT-OCT for OCT-based epiretinal membrane severity classification using hybrid CNN-transformer feature aggregation [8]. These studies indicate that retinal AI is moving towards practical models with better feature fusion, interpretability and computational efficiency.

However, most existing works mainly address the classification model. They do not solve the complete field-level workflow that includes patient positioning, automated image capture, local inference, display of results and remote ophthalmology review. Henceforth, the present work differs from existing studies by formulating retinal screening as a system-level engineering problem.

*Table 1. Comparison of related works with the proposed work*

| Paper No. | Disease Scope | Method | Limitation |
|---|---|---|---|
| [1] | Multi-label retinal disease | Retinal ViT | No robotic workflow |
| [2] | DR, HR, glaucoma, contrast disease | MobileNet + EfficientNet fusion | No edge deployment |
| [3] | 16 fundus diseases | DenseNet121 + XGBoost | Hospital dataset focus |
| [4] | DR, AMD, DME, CSR, MH | CNN + GNN + LLM | Multimodal data dependency |
| [5] | 9 retinal classes | YOLOv11 classifier | No acquisition automation |
| [6] | Fundus disease classification | IMF + LightCNN | Dataset-specific validation |
| [7] | Cataract, DR, glaucoma | Attention-based CNN/ViT | Limited disease range |
| [8] | ERM severity | DenseViT-OCT | OCT-only disease focus |
| Proposed work | Multi-disease retinal screening | Robot + edge AI + tele-review | Proposed framework |

**Key Contribution**

The primary contribution of this paper is the formulation of multi-disease retinal screening as a complete system-level engineering problem rather than only an image classification task. Recent works show good performance using CNNs, vision transformers, hybrid models and multimodal learning for retinal disease detection [1-8]. However, most of them mainly focus on model accuracy and do not fully address patient positioning, fundus image capture, local inference, result display and remote review together. Henceforth, this paper reviews recent open-access works and identifies the practical gap between high-performing retinal AI models and field-deployable screening systems.

**Method, Experiments and Results**

Recent studies from 2024 to 2026 were systematically reviewed with focus on fundus imaging, OCT imaging, multi-disease classification, multi-label prediction, CNN-transformer models, lightweight networks and multimodal AI. From the review done we can observe that retinal AI is shifting from single-disease detection to broader screening. Wang et al. used a retinal vision transformer for multi-label classification [1]. Khalid et al. proposed CAD-EYE using MobileNet and EfficientNet feature fusion and reported 98% accuracy on 65,871 fundus images [2]. Duan and Tu used 10,612 ultra-widefield fundus images for sixteen diseases and reported accuracy above 0.980 for common diseases [3]. Zedadra et al. then combined OCT, fundus images and clinical data using multimodal AI [4].

These works show the capability of CNNs are for local lesion extraction, while transformers support global retinal context learning. Suggesting, the hybrid and multimodal models improve disease coverage. However, most studies remain dataset-based and do not address field-level screening needs.

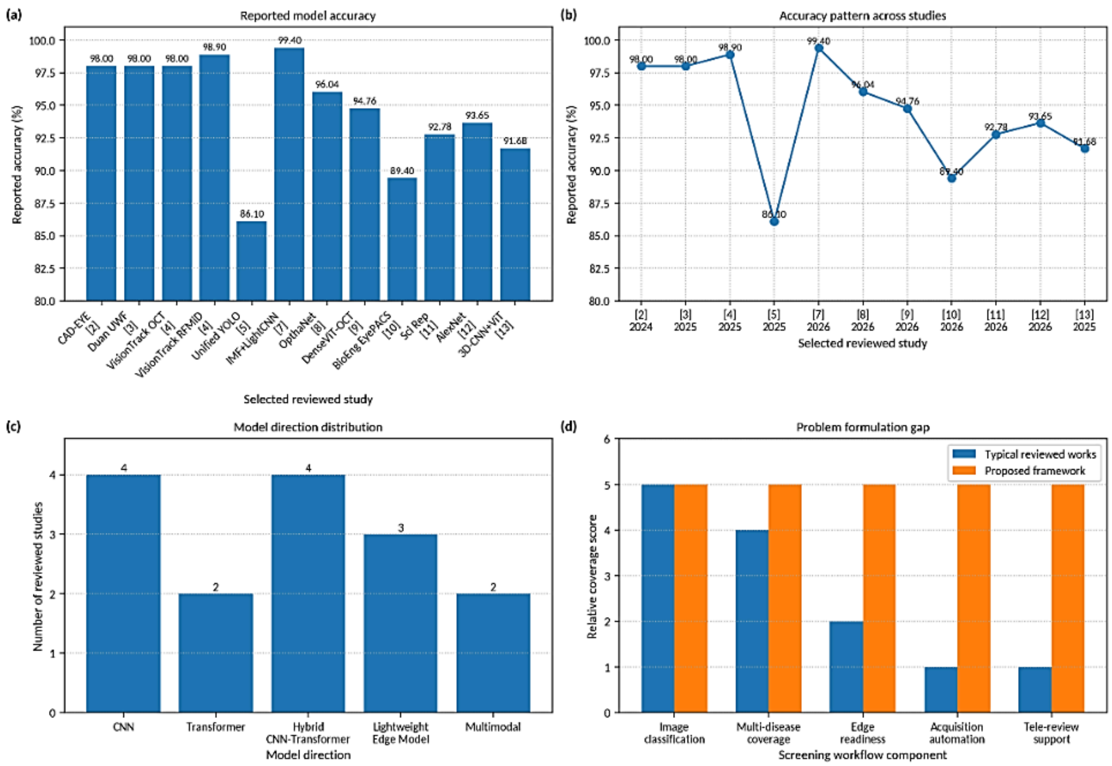


*Figure 1. Comparative analysis of reviewed retinal disease screening studies: (a) reported model accuracy across selected works, (b) accuracy trend pattern across recent studies, (c) distribution of model directions used in the reviewed literature, and (d) problem formulation gap between model-centric approaches and the proposed autonomous edge-AI retinal screening framework.*

**Discussions**

From the reviewed studies, we can observe that the retinal screening has moved from conventional CNN classification to transformer hybrid and multimodal, which improved the disease coverage and diagnostic performance in fundus and OCT-based screening. However, these works are model-centric and depend on standard evaluation parameters, whereas practical issues exist such as patient positioning, image quality variation, etc., and possibility for remote review is not fully addressed. Thus, a clear gap exists between high-performing retinal AI models and field-ready screening systems.

**Conclusions**

In this review, we have focused on multi-disease retinal screening for low-resource and field-level healthcare settings. We reviewed the recent studies from 2024 to 2026 and observed that CNNs, vision transformers, hybrid models, lightweight networks, and multimodal AI can support retinal disease detection with strong performance [1-12]. However, the most existing works are limited to dataset-based classification and do not cover the complete screening workflow. Therefore, this paper formulates autonomous retinal screening as a system-level problem involving patient guidance, non-mydriatic fundus image capture, edge-AI inference, result display and tele-ophthalmology support. Future work should validate the system using real-world datasets, edge-hardware testing and clinical trials.

## References


1. D. Wang, J. Lian and W. Jiao, "Multi-label classification of retinal disease via a novel vision transformer model", Frontiers in Neuroscience, vol. 17, Article 1290803, 2024. https://doi.org/10.3389/fnins.2023.1290803
2. M. Khalid, M. Z. Sajid, A. Youssef, N. A. Khan, M. F. Hamid and F. Abbas, "CAD-EYE: An automated system for multi-eye disease classification using feature fusion with deep learning models and fluorescence imaging for enhanced interpretability", Diagnostics,vol.14, no.23, Article 2679, 2024. https://doi.org/10.3390/diagnostics14232679
3. M.-M. Duan and X. Tu, "Deep learning-based classification of multiple fundus diseases using ultra-widefield images", Frontiers in Cell and Developmental Biology, vol. 13, Article 1630667, 2025. https://doi.org/10.3389/fcell.2025.1630667
4. A. Zedadra, M. Y. Salah-Salah, O. Zedadra and A. Guerrieri, "Multi-modal AI for multi-label retinal disease prediction using OCT and fundus images: A hybrid approach", Sensors, vol. 25, no. 14, Article 4492, 2025. https://doi.org/10.3390/s25144492
5. U. Şevik and O. Mutlu, "Automated multi-class classification of retinal pathologies: A deep learning approach to unified ophthalmic screening", Diagnostics, vol. 15, no. 21, Article 2745, 2025. https://doi.org/10.3390/diagnostics15212745
6. M. Shojaeinia, A. Hosseini, M. Naderi, B. Baloutch, M. S. Yekta, L. Akbarpour and H. Moghaddasi, "A comprehensive overview: Deep learning approaches to central serous chorioretinopathy diagnosis", BMC Ophthalmology, vol. 25, Article 549, 2025. https://doi.org/10.1186/s12886-025-04372-6
7. P. Kulkarni and K. S. Reddy, "Identification of retinal diseases using light convolutional neural networks and intrinsic mode function technique", Diagnostics, vol. 16, no. 5, Article 773, 2026. https://doi.org/10.3390/diagnostics16050773
8. S. R. Md. Nasif, S. M. Hasan Mahadi, Md. I. Mobin and M. F. Mridha, "OpthaNet: Attention-integrated architecture for high-precision multi-class ophthalmic image classification", Healthcare Technology Letters, vol. 13, Article e70067, 2026. https://doi.org/10.1049/htl2.70067
9. E. Yusufoğlu, S. T. A. Özçelik, O. Atila, N. H. Guldemir and A. Sengur, "DenseViT-OCT: A hybrid CNN-transformer architecture with multi-scale dense feature aggregation for automated epiretinal membrane severity classification", Tomography, vol. 12, no. 6, Article 76, 2026. https://doi.org/10.3390/tomography12060076
10. W. Abdelbaki, W. Bouchelligua, I. M. Nasir, S. Tehsin and H. Alshaya, "Transformer-based deep learning for population-scale retinal image screening of ophthalmic disorders", Bioengineering, vol. 13, no. 4, Article 377, 2026. https://doi.org/10.3390/bioengineering13040377
11. A. A. Hailu, E. A. Desta, A. Mulugeta, F. B. Adugna, A. O. Salau, L. A. Almaw and M. B. Fentahun, "Early retinal disease detection from fundus images using deep neural networks", Scientific

Reports, Article in Press, 2026.
https://doi.org/10.1038/s41598-026-42059-0

12. E.-S. M. Elkenawy, N. Khodadadi, K. S. Gaber, E. Khodadadi, A. A. Alhussan, D. S. Khafaga and M. M. Eid, "Automated retinal disease classification using deep learning and AlexNet with statistical models analysis", PLOS One, vol. 21, no. 1, Article e0338415, 2026. https://doi.org/10.1371/journal.pone.0338415